Descriptive Statistics and Normality Testing

Dr Daniel Boduszek
& Dr Katie Dhingra

QRM-TU
Quantitative Research Methods Training Unit

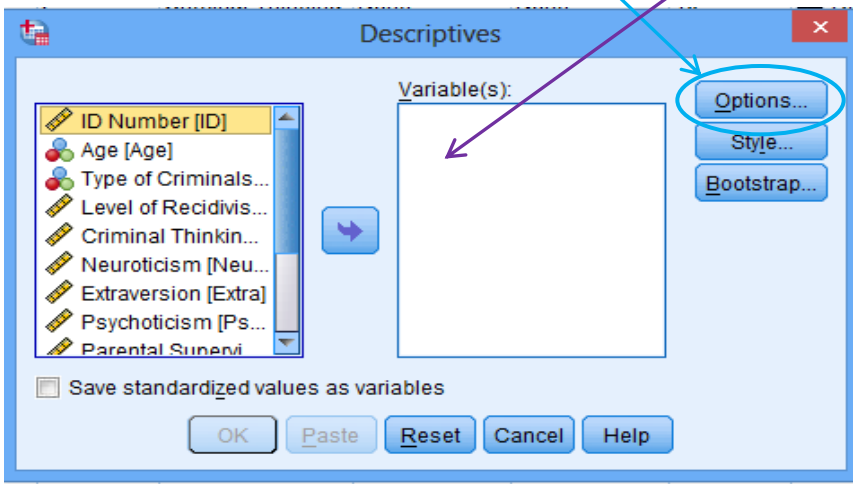
University of HUDDERSFIELD
Inspiring tomorrow's professionals

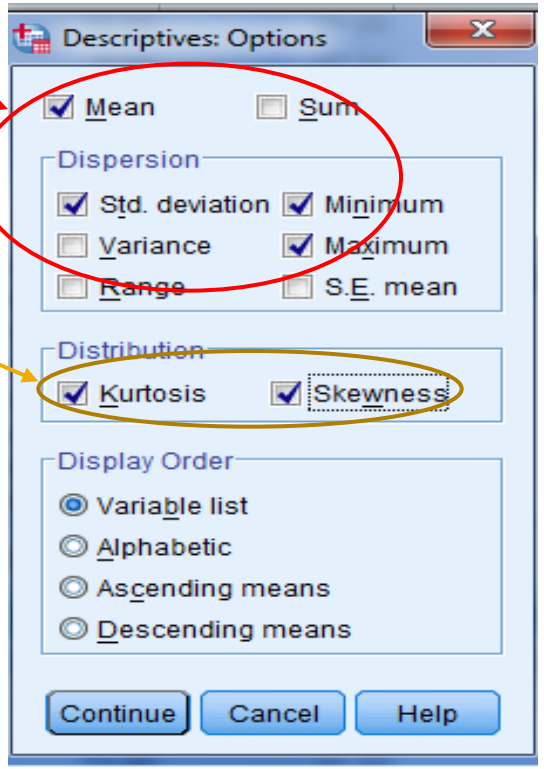
Introduction

- Descriptive Statistics help you to:
  - Describe the characteristics of your sample in the **Method section** of your thesis/publication
  - Check variables for any violation of the assumptions
  - Address specific research question

Descriptive Statistics

- **Continuous Variables**
  - **Analyze**, **Descriptive Statistics**, then **Descriptives**
  - Move variables that you want statistics for into the **Variable(s)** box
  - Click on **Options**. Tick **Mean**, **Std. deviation**, **Minimum** & **Maximum**
  - Click on **Skewness** & **Kurtosis**
  - Click **Continue** & **OK**





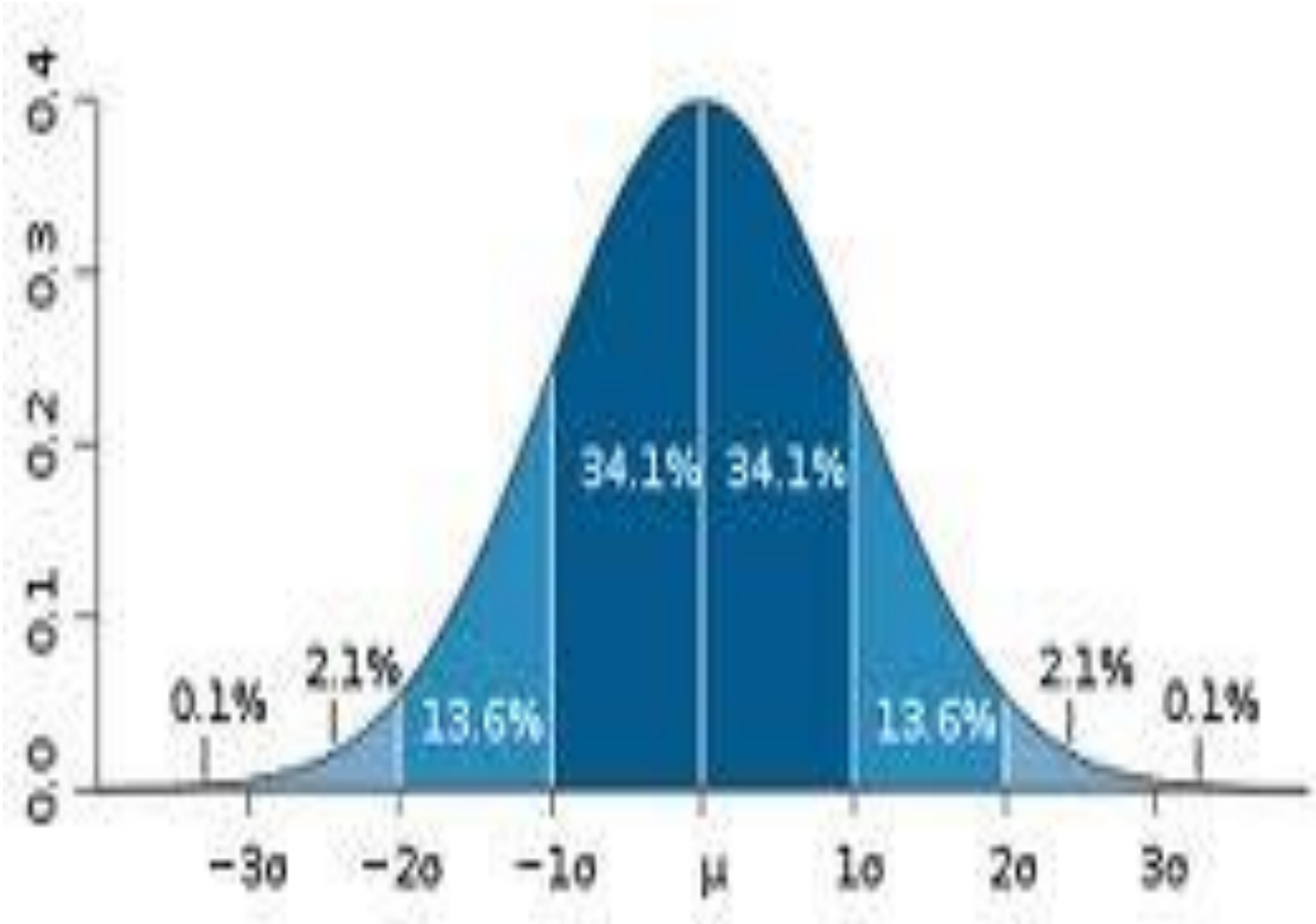
Descriptive Statistics

- ***N*** = number of participants
- **Min** & **Max** = range of the scores
- **Mean** = average score
- **Std. Dev** = Standard deviation (how much on average the individual values differ from the mean. The smaller the *SD* the less each score varies from the mean
- **Skewness** = provides an indication of the symmetry of the distribution
- **Kurtosis** = information on the "peakedness" of the distribution

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Criminal Thinking	89	10.00	43.00	29.1685	7.90747	-.240	.255	-.413	.506
Criminal Identity	89	-9.00	36.00	18.7303	8.93762	-1.094	.255	2.363	.506
Valid N (listwise)	89								

Normal Distribution



Normal Distribution

- Most data will be relatively normally distributed (bell-shaped)
- Normal distribution refers to a distribution of scores that normally occurs in a population
- Normal distributions indicate that the majority of individuals score in the middle range of collected data, with fewer people in the extreme high & low ends

Normal Distribution

- Normally distributed data is always what we would like to have
  - Inferential statistics are all designed to perform appropriately with normally distributed data.
- Using non-normally distributed data for inferential statistics is like using petrol in a diesel engine car.
  - Things tend to not to go well!

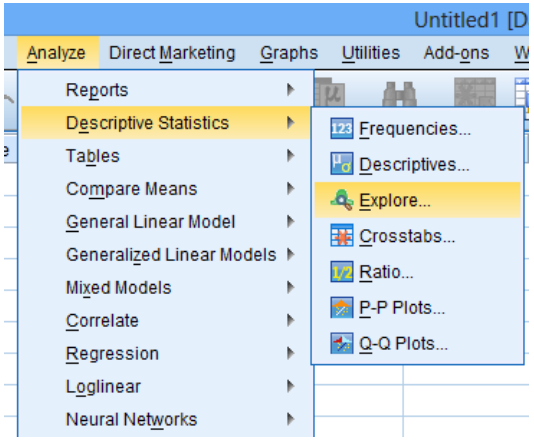
Normal Distribution

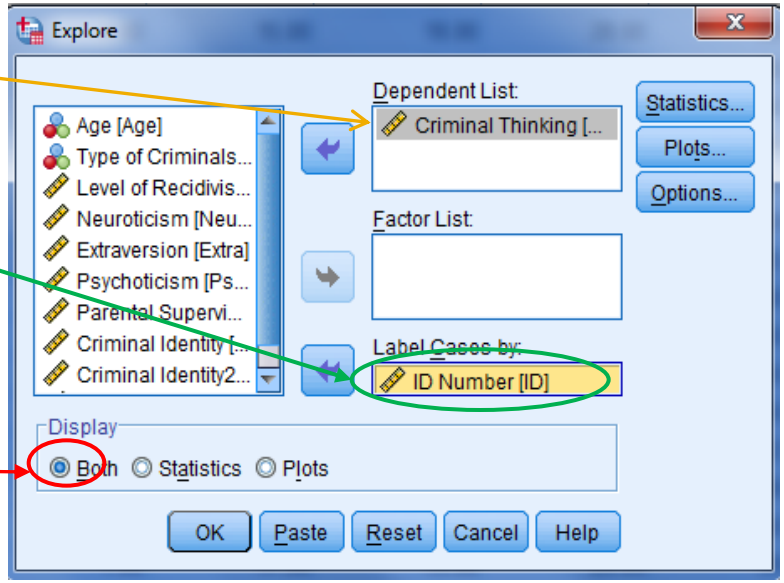
- There are four ways to tell if your data is normally distributed:

1. Inspect your mean, mode, and median scores.
2. Inspect a histogram and fit a normal curve to visually determine it's normality.
3. Look at the skewness and kurtosis along with their standard errors.
4. Perform a **statistical test**.

Assessing Normality

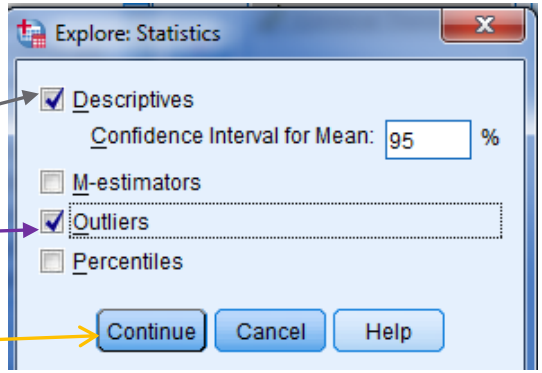
- **SPSS procedure**
  - **Analyze, Descriptive Statistics,** then **Explore**
  - Click on the variable and move it into **Dependent List** box
  - In the **Label Cases by** – put ID variable
  - In the **Display** section – select **Both**

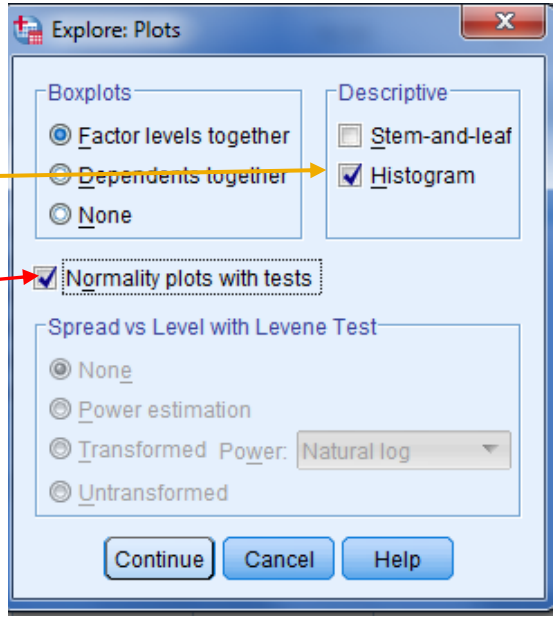




Assessing Normality

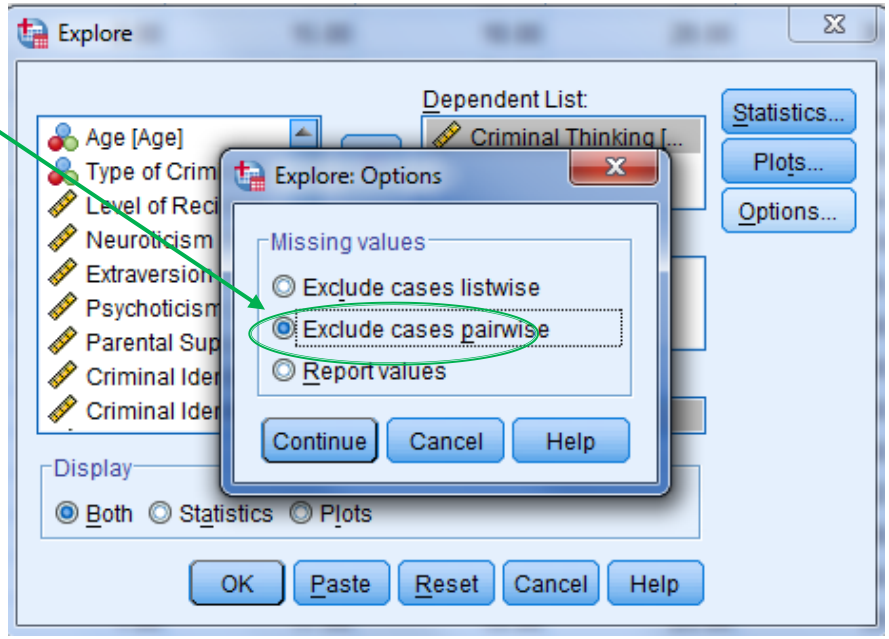
- Click on **Statistics**, select **Descriptives** & **Outliers**
- **Continue**
- Click on **Plots**.
  - Click **Histogram**
  - Click **Normality plots with tests**
  - **Continue**





Assessing Normality

- Click on the **Options**
- **Exclude cases pairwise** (excludes the cases only if they are missing the data required for the specific analysis)
- **Continue** & **OK**



Assessing Normality

Descriptives

			Statistic	Std. Error
Criminal Thinking	Mean		29.1685	.83819
	95% Confidence Interval for Mean	Lower Bound	27.5028	
		Upper Bound	30.8343	
	5% Trimmed Mean		29.3733	
	Median		28.0000	
	Variance		62.528	
	Std. Deviation		7.90747	
	Minimum		10.00	
	Maximum		43.00	
	Range		33.00	
	Interquartile Range		11.00	
	Skewness		-.240	.255
	Kurtosis		-.413	.506

- If a skew or kurtosis result is more than twice its standard error than you may have a problem with normality.
- On the other hand, skew is not considered problematic unless its value is greater than +/- 1.
- If skewness is OK then there is no need to worry about kurtosis.

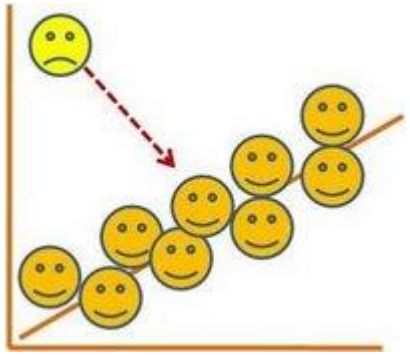
- In a normal distribution the mean, mode, & median will all be identical!

Assessing Normality

- **Extreme Values table** – ID values of the most extreme cases

Extreme Values

			Case Number	ID Number	Value
Criminal Thinking	Highest	1	20	20	43.00
		2	88	88	43.00
		3	19	19	42.00
		4	87	87	42.00
		5	89	89	42.00
	Lowest	1	5	5	10.00
		2	44	44	12.00
		3	35	35	13.00
		4	33	33	13.00
		5	42	42	14.00



Assessing Normality

- **Test of Normality** table – non-significant value (Sig > .05) indicates normality

Tests of Normality

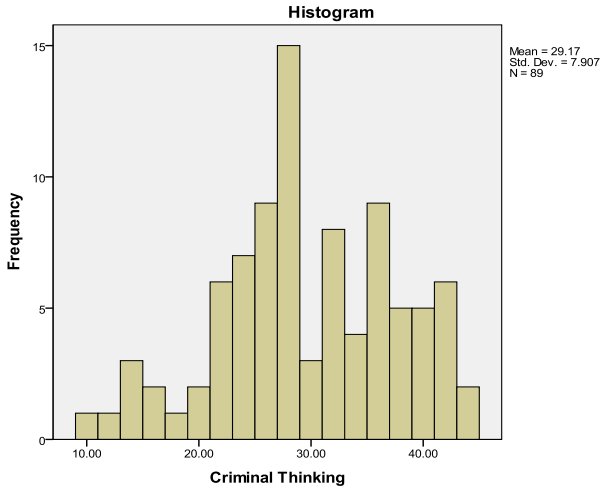
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Criminal Thinking	.087	89	.094	.975	89	.080

a. Lilliefors Significance Correction

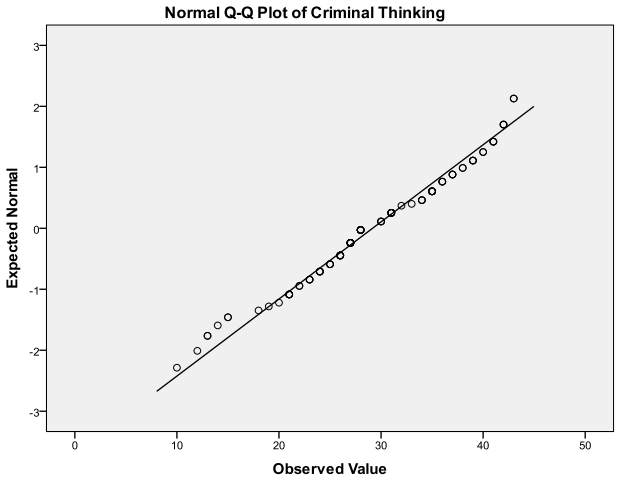
Less than 40 cases in your data set

Assessing Normality

- **Histogram** – check the shape of distribution



- **Normal Q-Q Plot**
  - Reasonably straight line suggests a normal distribution



Normality

- Very sensitive to both small & large data sets.
  - Small – not enough power and inability to detect any variation from normality – thus non-normally distributed data can be erroneously deemed to be normally distributed.
  - Large – power to detect even minute deviations from normality – thus normally distributed data can be erroneously deemed to be non-normally distributed.

Normality

- Inspect the test of normality results but also inspect your histograms & if your data is 'normal enough' you are good to go!
  - Provided you have a large enough sample
- With a big enough sample (> than 100) minor violations of normality are not a problem – using clever witchcraft SPSS is able to take care of these violations.

Thank you for your time!